EXHIBIT 1

Evidence of Fraud in Academic Articles Authored By Francesca Gino

1. Introduction

This report reflects a collaboration among a group of anonymous researchers. A small number of individuals raised concerns to us and asked for our involvement in trying to reconcile those concerns. In collaboration, we have collectively tried to identify some of the biggest issues. Rather than each individual making public their concerns, we have elected to present this evidence to Harvard University so that its investigators can consider the case while giving full opportunity for Professor Gino to explain apparent anomalies.

We report direct evidence of data tampering in four different datasets from four different published articles. We focus on those because they appear the most unambiguous. We have strong suspicions about some her published data going as far back as 2008 (when she was a post-doc at Carnegie Mellon University), but the most direct evidence is included it in this report.

Indeed, we should be clear that neither this report, nor our investigation, are exhaustive. We have not analyzed, or even read, the majority of Professor Gino's published articles. If the Harvard University investigators determine that there is sufficient evidence in these four studies, it would certainly be worth considering others as well.

Finally, although the evidence can, in most of these cases, rule out malfeasance by co-authors, it cannot definitively rule in malfeasance by Professor Gino. It may be that some research assistant or otherwise unnamed person/people was/were responsible for producing these anomalies.

2. Overview

While the substantive research questions, manipulations, and dependent variables across the four papers are quite different, the anomalies we uncovered share a few commonalities worth keeping in mind as one examines the evidence. The commonalities suggest *imperfect* data tampering; that is to say, the datasets have features consistent only with tampering, but also features that could have been potentially detected and eliminated by the person doing the tampering.

One common anomaly consists of datasets that are sorted, but sorted imperfectly. Imperfect sorting left a trace of rows that were moved and/or values that were changed. For example, imagine a dataset sorted by participant ID, but in which some observations are out of sequence, say IDs being 1, 2, 3, 4, 81, 82**,** 5, 6, 7. If the out-of-order rows of data (e.g., those with IDs **81** and **82**) exhibit extraordinarily large effect sizes – at the extreme, effect sizes that produce the overall effect in its entirety – then that would represent fairly strong evidence of data tampering. We find that in two of these cases.

Another common anomaly consists of answers provided by participants that are inconsistent with the question being asked (e.g., participants answering "Harvard" to the question of how many years they have spent at school), or with other values in the dataset (e.g., participants indicating they felt maximally disgusted with a networking event and then describing that same networking event with words such as "exciting" and "great"). Those anomalous observations, again, show extraordinarily large effects consistent with the researcher's hypothesis.

3. Case #1: Study 1 of Shu, Mazar, Gino, Ariely, and Bazerman (2012)

In this paper, the authors present three studies suggesting that "signing before—rather than after—the opportunity to cheat makes ethics salient when they are needed most and significantly reduces dishonesty" (page 15197).

Here we focus on Experiment 1, which was run at the University of North Carolina (UNC). Our understanding is that Gino supervised the execution of this experiment, and analyzed the data, but perhaps it is worth checking with co-authors to make sure. It is possible that an RA assisted Gino (e.g., Jennifer Fink is thanked in the acknowledgements; she has an online presence as a life coach, making it easy to contact her if deemed appropriate by those investigating these matters).

3.1 Procedure

In Experiment 1, 101 participants first completed a math puzzles task. "Participants were told that they would have 5 min to find two numbers in each puzzle that summed to 10. For each pair of numbers correctly identified, they would receive $1, for a maximum payment of $20. Once the 5 min were over, the experimenter asked participants to count the number of correctly solved puzzles, note that number on the [anonymized] collection slip, and then submit both the test sheet and the collection slip to the experimenter." Note that participants had the ability and incentive to cheat on this task, by simply overreporting the number of puzzles that they solved on that collection slip.

After this task, participants filled out a one-page "tax return form." On that form, participants reported both how much money they had earned from the math puzzles task, as well as "how many minutes it took them to travel to the laboratory, and the cost of their commute. These expenses were 'credited' to their posttax earnings from the [math puzzles] task to compute their final payment." Thus, participants were motivated not only to overreport their math puzzle task performance, but also to overreport the cost of their commute.

The critical intervention in this study involved the format of the "tax return form." Participants were randomly assigned to one of three conditions. In the *sign-at-the-top* condition, participants had to sign at the top of the page, under a statement that read, "I declare that I carefully examined this return and that to the best of my knowledge and belief it is correct and complete." In the *sign-at-the-bottom* condition, participants instead signed at the bottom of the page. And in the *control* condition, participants did not sign the form at all.

In sum, this experiment featured one independent variable – the placement of the signature on the tax return form – and two dependent variables – (1) how much participants cheated on the math puzzles task[1] and (2) how many expenses they claimed for their commute on the tax-return form.

3.2 Reported Results

Participants in the *sign-at-the-bottom* condition overclaimed fewer correct solutions (M=.77) than those in the *sign-at-the-bottom* condition (M=3.94), $p < .001$. Similarly, they claimed lower commuting expenses (M=$5.27, vs M=$9.62, $p < .01$). These are very big effects: Signing at the bottom vs. top quadrupled cheating on the math task, and doubled cheating on claimed commuting expense.

3.3 Anomaly: Out-of-Order Observations In The Dataset

[1] Because of a clever design feature of the math puzzles task, the researchers could link participants' reported math puzzle performance to their actual math puzzle performance. Thus, the researchers could compare how many math puzzles participants reported solving to how many puzzles they actually solved.

We retrieved the dataset for Experiment 1 from the OSF, where, since 2020, it has been publicly posted (https://osf.io/4b7mu/).

The posted dataset seems to be sorted by two columns, first by a column called "Cond", indicating participants' condition assignment (0 = control; 1 = sign-at-the-top; 2 = sign-at-the-bottom), and then by a column called "P#", indicating a Participant ID number assigned by the experimenter.

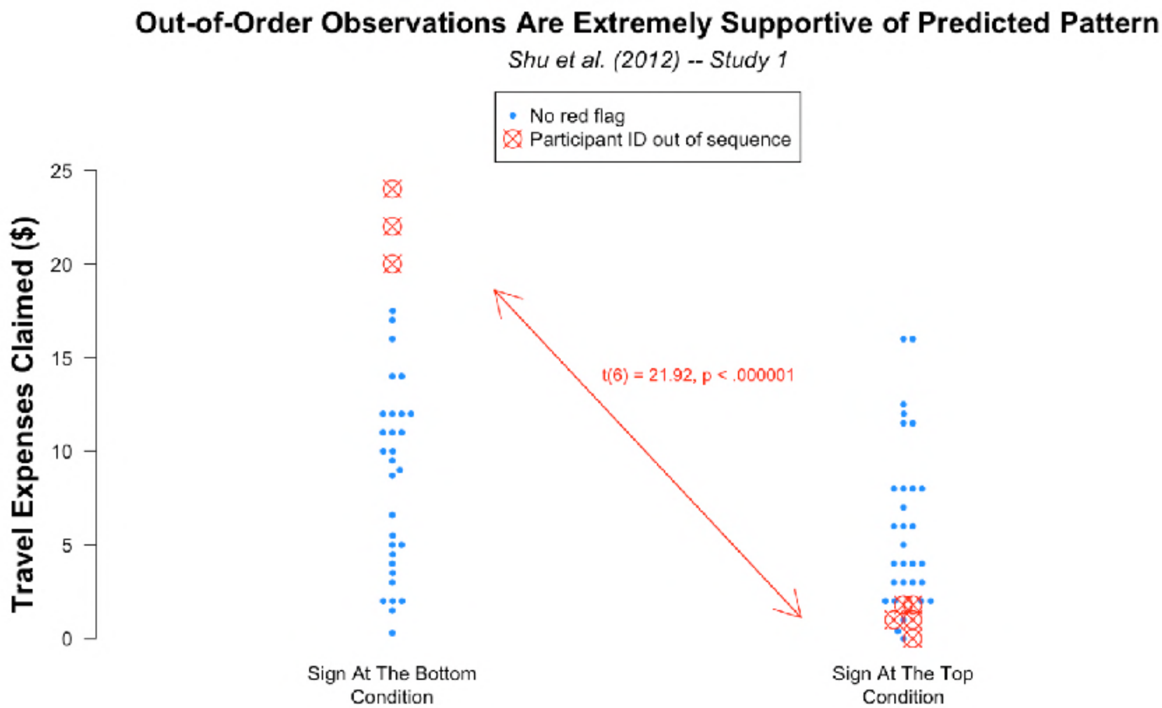
For example, this is a screenshot of a few dozen observations from the sign-at-the-top and sign-at-the-bottom condition. You can see that within each condition the data are *almost* perfectly sorted by Participant ID (the first column on the left). However, we have highlighted eight observations that are out of order:[2]

	A	B	C	D	E	F	G	H	I
1	**P#**	**Cond**	**Stude**	**Major**	**CS3**	**Male**	**Age**	**#B**	**$B**
47	35	1	1	Journalism	3	1	19	12	12
48	37	1	1	Economics	4	0	21	9	9
49	40	1	1	Political Science	5	1	29	15	15
50	42	1	1	Political Science	3	0	20	7	7
51	46	1	1	Political Science	4	0	21	12	12
52	49	1	1	English	4	1	21	9	9
53	49	1	1	English	4	1	21	7	7
54	55	1	1	Biology	4	1	21	12	12
55	58	1	1	Environmental Sciences	3	0	20	10	10
56	61	1	1	Nursing	3	0	20	15	15
57	63	1	0	NA		0	22	12	12
58	68	1	1	Business	3	1	20	16	16
59	70	1	1	Chemistry	4	0	21	11	11
60	73	1	1	Chemistry	5	0	20	16	16
61	76	1	1	Chemistry	2	1	19	15	15
62	80	1	1	Nursing	4	0	21	15	15
63	82	1	1	Economics	4	1	21	9	9
64	85	1	1	Psychology	4	0	20	5	5
65	88	1	1	Chemistry	3	0	20	13	13
66	95	1	1	Math Education	3	1	22	13	13
67	51	1	0	NA	0	0	52	4	4
68	12	1	1	Psychology	3	0	20	13	13
69	101	1	0	Business	3	1	20	6	6
70	7	2	0	Political Science	5	1	22	17	17
71	91	2	1	Japanese	2	1	20	17	17
72	52	2	0	NA	5	0	22	8	8
73	5	2	1	Biology/Psychology	2	0	18	16	16
74	8	2	1	Communication Studies	4	0	22	15	15
75	13	2	1	Chemistry	4	0	20	18	18
76	17	2	1	Communications	4	0	21	14	14
77	18	2	1	Communications	4	1	22	13	13
78	22	2	0			0	23	13	13
79	26	2	0			0	47	6	6
80	27	2	1	Mathematics - Sociology	3	1	19	18	18

Participant ID 49 appears twice in the dataset, with identical demographic variables. In addition, Participants 51, 12, 101 are out of order in Condition 1, and Participants 7, 91, and 52 are out of order in Condition 2. We see this as a red flag because, to our knowledge, there is no way to sort the data in a way that achieves this ordering. It suggests that observations must have been moved around (or duplicated), manually, perhaps to alter a participant's condition assignment in a way that achieves the desired result.

A deeper dive into the data of these eight participants provides support for this form of data tampering. The figure below shows a "Bee Swarm" plot, which depicts each observation in the dataset, separately for each experimental condition. The plot depicts one of the cheating measures, the amount of money participants claimed in travel expenses. Every "normal", in-sequence observation is represented as a blue dot, whereas the eight out-of-sequence observations are represented as red X's.

[2] There is one additional out-of-order observation in the control condition (not shown). But for simplicity we focus our analyses on the comparison between the sign-at-the-bottom and sign-at-the-top conditions. That one out-of-order control condition observation scored highly on overreporting math puzzles, with a score of 4 (the median is 1), and low on travel expenses claimed ($1).



In the sign-at-the-bottom condition, the authors predicted expenses to be high, and indeed the three out-of-sequence observations in this condition are the very highest. In the sign-at-the-top condition, the authors predicted expenses to be low, and indeed the five out-of-sequence observations in this condition were all among the very lowest. As shown in the plot, the condition difference between just these eight observations on this dependent variable is very highly significant; it would occur by chance less than 1 in a million times.[3] We have been unable to generate a benign explanation for this pattern.

A similar effect emerges when analyzing the other dependent variable, the overreporting of the number of math puzzles solved. The five out-of-sequence observations in the sign-at-the-top condition, predicted to be low, are all equal to zero, the lowest value observed in the dataset. The three out-of-sequence observations in the sign-at-the-bottom condition, predicted to be high, were all greater than zero: 2, 6, and 7. The condition difference between these eight observations on this dependent variable was again highly significant, even with so few observations: $t(6) = 4.48$, $p = .004$.[4]

In sum, there are eight observations that are out of order in this dataset, and to our knowledge no sorting function can account for their placement. This suggests to us that these eight observations may have been altered to produce the desired effect. Supporting that contention, those eight observations play a sizable

[3] This p-value (probably correctly) assumes that there are truly no differences between conditions. We ran 1 million simulations that examined what this p-value would be if we instead very conservatively assumed that the condition differences are exactly as large as what was observed in the data. In each simulation, we drew five observations at random from the sign-at-the-top condition and three observations at random from the sign-at-the-bottom conditions (without replacement), mirroring the number of flagged observations we observed in each condition in the data. We then conducted a t-test to analyze the condition difference between those observations. We observed a t-value as large as what we observed for the flagged observations (21.92) only 10 times in those 1 million simulations, suggesting a p-value of 1 in 100,000. Thus, even when we assume that the true condition differences are exactly as large as they are in the observed dataset, there is only an extremely small chance of finding such a large condition difference among a *randomly* selected subset of eight observations.

[4] Using the same conservative approach described in the previous footnote, the p-value is .065.

role in producing the published effect in Study 1, as all eight observations have values on the dependent variables that are extremely consistent with the authors' hypothesis.

Before moving on, we should be clear that we do not believe that these eight observations are necessarily the *only* ones that may have been tampered with. Rather, they may be a mere subset, identifiable only because the person tampering with the data neglected to re-sort the dataset. We cannot identify every instance of fraud. We can only identify it when those doing the tampering leave observable traces of what they have done.

4. Case #2: Study 4 of Gino, Kouchaki, and Galinsky (2015)

In this paper, the authors present five studies indicating that "experiencing inauthenticity, compared with authenticity, consistently led participants to feel more immoral and impure. The link from inauthenticity to feeling immoral produced an increased desire among participants to cleanse themselves and to engage in moral compensation by behaving prosocially" (p. 983).

Here we focus on Experiment 4, which was run at Harvard University. Participants' responses to a question about their "class year" in the dataset indicate that the study was run no earlier than Fall of 2014, as seniors reported being in the Class of 2015, juniors in the Class of 2016, and so on. Although the second author of this paper, Maryam Kouchaki, was a postdoctoral researcher at Harvard for two years, her cv indicates that she began her job as an Assistant Professor at Northwestern in 2014, making it very unlikely that she was still at Harvard when this study was conducted and analyzed. In addition, the data file and methods write-up posted on the OSF website were uploaded by Gino, and the properties of those files indicate that she created them. Thus, it is most likely that this study was run/supervised and analyzed by Gino. With all of that said, that can only be verified by Harvard University.

4.1 Procedure [5]

Harvard students (N = 491) came into a lab and were first "asked to confirm that they were college students at Harvard." They were then "asked for their opinion [on] whether or not difficulty ratings should be a part of the Q guide (in which all Harvard courses are rated and reviewed by students who have taken them in the past)." Participants were then "asked for their age, gender, and year in school. They were then told that their first task was to write an essay on a current topic."

During the essay task, participants were randomly assigned to one of three conditions. One-third were asked to write an essay in support of their opinion about including difficulty ratings in the Q guide (the pro-attitudinal condition), and two-thirds were asked to write an essay *against* their opinion that that issue (the counter-attitudinal conditions). The two-thirds asked to write a counter-attitudinal essay were randomly assigned to one of two conditions, involving how much choice they had as to whether to write such an essay: low-choice vs. high-choice. Thus, the three essay conditions were (1) pro-attitudinal, (2) counter-attitudinal (low-choice), and (3) counter-attitudinal (high-choice).

After writing the essay, "participants received a list of products and indicated how desirable they found them to be . . . We averaged ratings of the five cleansing products to create one aggregate measure."

The authors hypothesized that "participants would express a greater desire for cleanliness whenever they wrote essays that were not consistent with their internal beliefs, regardless of their perceived level of choice." That is, they predicted that participants' preference for cleaning products would increase after

[5] This section frequently quotes directly from the introduction and methods of this study, as written up in Gino et al. (2015, p. 991-992).

writing a counter-attitudinal essay, regardless of whether they did so under conditions of low choice or high choice.

4.2 Results

Consistent with the authors' hypotheses, participants were less desirous of cleaning products in the pro-attitudinal condition ($M=3.72$, $SD=1.33$), compared to both the counter-attitudinal (high choice) condition, ($M=4.18$, $SD=1.51$), $p=.012$, and the counter-attitudinal (low choice) condition ($M=4.34$, $SD=1.44$), $p < .001$.

4.3 The Anomaly: Strange Demographic Responses

As mentioned above, students in this study were asked to report their demographics. Here is a screenshot of the posted original materials, indicating exactly what they were asked and how:

4. Your age: _______

5. Your gender
- Male
- Female
- Other (please indicate)

6. Year in School: ________

We retrieved the data from the OSF (https://osf.io/sd76g), where it has been posted since 2015. The anomaly in this dataset involves how some students answered Question #6: "Year in School."

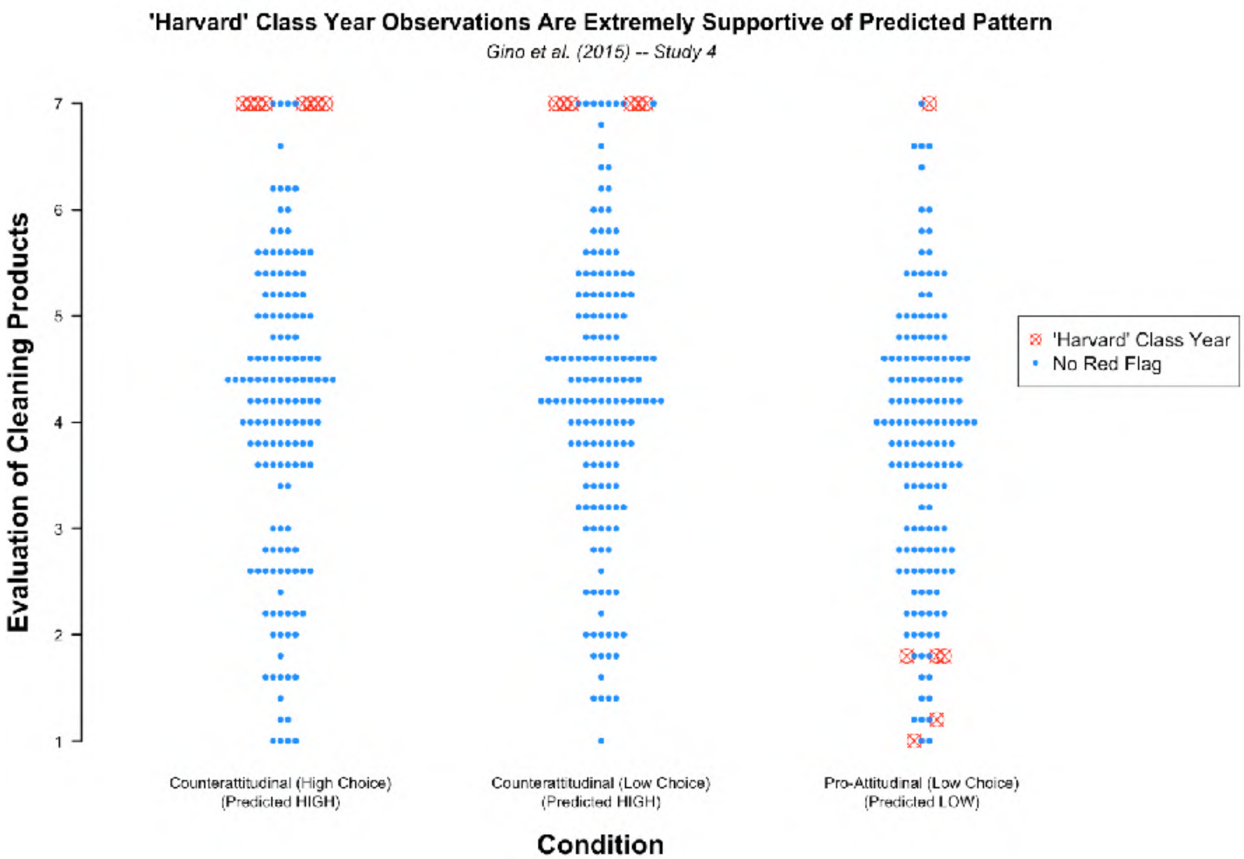
The screenshot below shows a portion of the dataset. In the "yearSchool" column, you can see that students approach this "Year in School" question in a number of different ways. For example, a junior might write "junior", or "2016" or "class of 2016" or "3" (to signify that they are in their third year). All of these responses are reasonable.

A less reasonable response is "Harvard", an incorrect answer to the question. It is difficult to imagine many students independently making this highly idiosyncratic mistake. Nevertheless, the data file indicates that 20 students did so. Moreover, and making things even more peculiar, those students' responses are very close to one another, all within 35 rows (450 through 484) in the posted dataset:

1	instr	college_st	inFavor_ir	strongOp	age	male	gender_T	yearSchool	condition	i
443	1	1	1	7	19	1		Sophomore	ProAttitudinal	
444	1	1	1	7	20	1		Junior	No_Choice	
445	1	1	1	6	19	0		sophomore	High_Choice	
446	1	1	1	6	20	1		Junior	ProAttitudinal	
447	1	1	1	7	21	1		Senior (Class of 201	No_Choice	
448	1	1	1	5	22	1		Senior	High_Choice	
449	1	1	1	5	21	1		Senior	ProAttitudinal	
450	1	1	1	7	23	0		harvard	No_Choice	
451	1	1	1	4	21	0		2015	High_Choice	
452	1	1	1	7	20	1		Junior	No_Choice	
453	1	1	1	7	18	0		Sophomore	ProAttitudinal	
454	1	1	0	7	25	0		Harvard	High_Choice	
455	1	1	0	7	25	0		Harvard	ProAttitudinal	
456	1	1	1	7	22	1		Harvard	ProAttitudinal	
457	1	1	0	7	24	0		Harvard	High_Choice	
458	1	1	1	7	22	0		Harvard	High_Choice	
459	1	1	0	7	25	0		Harvard	No_Choice	
460	1	1	1	7	23	1		Harvard	ProAttitudinal	
461	1	1	0	7	25	0		Harvard	High_Choice	
462	1	1	0	6	25	1		4	No_Choice	
463	1	1	0	7	24	0		Harvard	No_Choice	
464	1	1	1	5	23	0		1	High_Choice	
465	1	1	1	4	19	0		Sophomore	No_Choice	
466	1	1	1	6	28	1		5	High_Choice	
467	1	1	1	6	22	1		Senior	ProAttitudinal	
468	1	1	1	6	20	0		Junior	High_Choice	
469	1	1	1	5	23	1		2015	High_Choice	
470	1	1	1	6	22	1		Senior	No_Choice	
471	1	1	1	6	22	0		2015/Senior	ProAttitudinal	
472	1	1	1	6	36	1		2010	High_Choice	
473	1	1	1	7	25	0		Harvard	ProAttitudinal	
474	1	1	0	5	25	0		Harvard	High_Choice	
475	1	1	1	7	22	1		Harvard	No_Choice	
476	1	1	1	7	23	1		Harvard	High_Choice	
477	1	1	0	7	25	0		Harvard	ProAttitudinal	
478	1	1	0	7	26	1		Harvard	No_Choice	
479	1	1	1	6	20	0		2013	No_Choice	
480	1	1	0	6	21	0		2012	ProAttitudinal	
481	1	1	1	7	24	1		Harvard	High_Choice	
482	1	1	1	7	27	0		Harvard	ProAttitudinal	
483	1	1	1	7	25	1		Harvard	High_Choice	
484	1	1	1	7	27	0		Harvard	No_Choice	
485	1	1	1	7	26	1		4	High_Choice	
486	1	1	0	6	22	0		2012	High_Choice	
487	1	1	1	6	20	1		2013	No_Choice	

This is a red flag, for it could indicate that someone had copy-pasted rows of data, without noticing that it resulted in an implausible number of students providing the same strange and erroneous answer to a straightforward question.

If these peculiar observations were indeed tampered with, then we should see that students who answered "Harvard" were especially likely to confirm the authors' hypothesis. To see this, we again present a Bee Swarm plot, which depicts each observation in the dataset, separately for each experimental condition. The plot depicts the key dependent variable, participants' average ratings of how much they desired five cleaning products. Every "normal", in-sequence observation is again represented as a blue dot, whereas the 20 "Harvard" observations are represented as red X's:



Here you can see that in the two counter-attitudinal conditions, which were predicted to induce a desire for cleaning products and thus higher values on y-axis, every "Harvard" observation has the highest possible average value (i.e., a 7.0). Conversely, in the pro-attitudinal condition, which was predicted to induce a lower desire for cleaning products, every "Harvard" observation is associated with a low value, except for one (which itself happens to be the only one associated with a lowercase "harvard").

The difference between the Pro-Attitudinal and Counter-attitudinal conditions for just these 20 observations is highly significant, with a p-value indicating that it would occur by chance less than one in a million times: $t(18) = 7.84$, $p < .000001$.[6]

As in Case #1, this is very much consistent with the possibility that these "Harvard" observations were altered to produce the desired effect.

5. Case #3: Study 3a of Gino, Kouchaki, and Casciaro (2020)

In this paper, the authors present six studies examining "how self-regulatory focus, whether promotion or prevention, affects people's experience of and outcomes from networking. [They] find that a promotion

[6] We also took the same conservative approach described in Footnote 3. In 1 million simulations, we observed a t-value as large as 7.84 only six times. Thus, under the assumption that the between condition difference between the counter-attitudinal vs. pro-attitudinal condition was identical to what was observed in the data, we would expect a "Harvard" class year pattern that is so highly predictive of the authors' result to emerge by chance only about 1 in 167,000 times.

focus, as compared to a prevention focus or a control condition, is beneficial to professional networking, as it lowers feelings of moral impurity from instrumental networking" (p. 1221).

Here we focus on Experiment 3a, which was run online (using mTurk participants). We believe it was conducted and analyzed by Gino because the materials posted on the OSF list "Qualtrics" as the creator of the file and "Francesca Gino" as the last person to save it. Thus, it is very likely that this was run through her Qualtrics account, which it turn makes it very likely that she analyzed the data. Only Harvard University can verify that fact.

5.1 Procedure. [7]

In Study 3a, 599 working adults recruited through MTurk first completed a writing task, during which they were randomly assigned to one of three conditions. Participants in the promotion-focus condition wrote about a current hope or aspiration, participants in the prevention-focus condition wrote about a current duty or obligation, and participants in the control condition wrote about what they do on a typical evening.

Participants then read a story in which they imagined "being invited to attend an event during which they socialized with other people. In the story the main character was described as 'actively and intentionally making professional connections with the belief that connections are important for future professional effectiveness.'"

Participants were then asked "to report how they felt at that moment, by indicating the extent to which they felt . . . dirty, inauthentic, and impure, ashamed, wrong, unnatural, and tainted." They did this using a scale ranging from 1 = not at all to 7 = very much. Participants then were asked to reflect on their previous writing task for 1-2 minutes, and to then "write a few words that came to mind regarding the story before proceeding to the next task." Participants completed other measures after that, but our focus is going to be on (1) the 7-item measure of moral impurity and (2) the words that participants wrote about the networking task, and so we won't describe those details here.

5.2 Results

As predicted, average scores on the 7-item moral impurity measure differed significantly across conditions, $F(2, 596) = 17.69$, $p < .0000001$. Ratings of moral impurity were significantly higher in the prevention-focus condition than in the control condition, which was in turn significantly higher than in the promotion-focus condition.

5.3 Direct Evidence of Tampering

It is useful to begin by looking at the Study 3a dataset. The screenshot below shows data for 22 participants (1 per row) for the key variables in this dataset:

[7] This section frequently quotes directly from the methods of this study, as written up in Gino et al. (2020, p. 1229-1230).

Ratings **of feeling cheap, dirty etc during networking event**

Words **describing networking event**

	E	F	G	H	I	J	K	L	M	
1	consent	essay	MI1	MI2	MI3	MI4	MI5	MI6	MI7	words2_cond
531	1	I relax	1	1	1	1	1	1	1	socializing,party,impression,connections,work
532	1	I ought to	2	2	2	2	2	2	2	success, happy, promotion, networking, impressive, connections
533	1	Retire muc	1	1	1	1	1	1	1	making the money for my dream
534	1	My husbar	1	1	2	1	1	1	1	Interaction, first impresions, career, goals, shmoozing, socializing
535	1	I would lik	1	1	2	1	1	1	1	Making connections to help myself
536	1	I walk hom	1	1	1	1	1	1	1	Wow, liar, false, delusional, braggart
537	1	I usually sp	2	1	2	3	2	3	2	I felt very happy and excited.
538	1	I cook dinr	1	1	2	1	1	3	1	proud, accomplished, social, quick-witted, happy
539	1	I want to r	1	1	1	1	1	1	1	dirty,fake,cheap,butt kisser,not a good person
540	1	I usually gc	1	1	3	1	1	3	1	Stressful, unique, performative, important, judgmental, impactful,
541	1	I would idc	1	1	1	1	1	1	1	i am glad that i made a good impression on everyone
542	1	well i norn	1	1	1	1	1	1	1	that i was very wise to use the party to connect with co workers
543	1	I would lik	2	3	3	3	2	3	3	contacts, mingling, smiling, fake, corporate, parties
544	1	I would lik	3	2	6	5	2	6	3	fake, boring, exhausting, tiring, dreadful
545	1	I ought to	2	2	2	1	1	1	1	smart,manipulative, lucky, future, sly
546	1	i would lik	1	1	1	1	1	1	1	important, friendly, high class, important, exciting
547	1	I would lov	1	1	1	1	1	1	1	Lucky, smart, determined, sharp, social
548	1	My aspirat	1	1	1	1	1	1	1	Happiness, joy, content, excitement
549	1	I head hon	1	1	1	1	1	1	1	Trying to succeed, keeping my position at this job
550	NA	I currently	2	2	2	2	2	2	2	go getter, intelligent, goal oriented, strong, not afraid, not shy
551	1	I would lik	5	5	5	5	5	5	5	inauthentic, oppressive, false, awkward, corrupt
552	1	One aspira	1	1	1	1	1	1	1	Proud, anxiety, pleased, cheerful, supportive

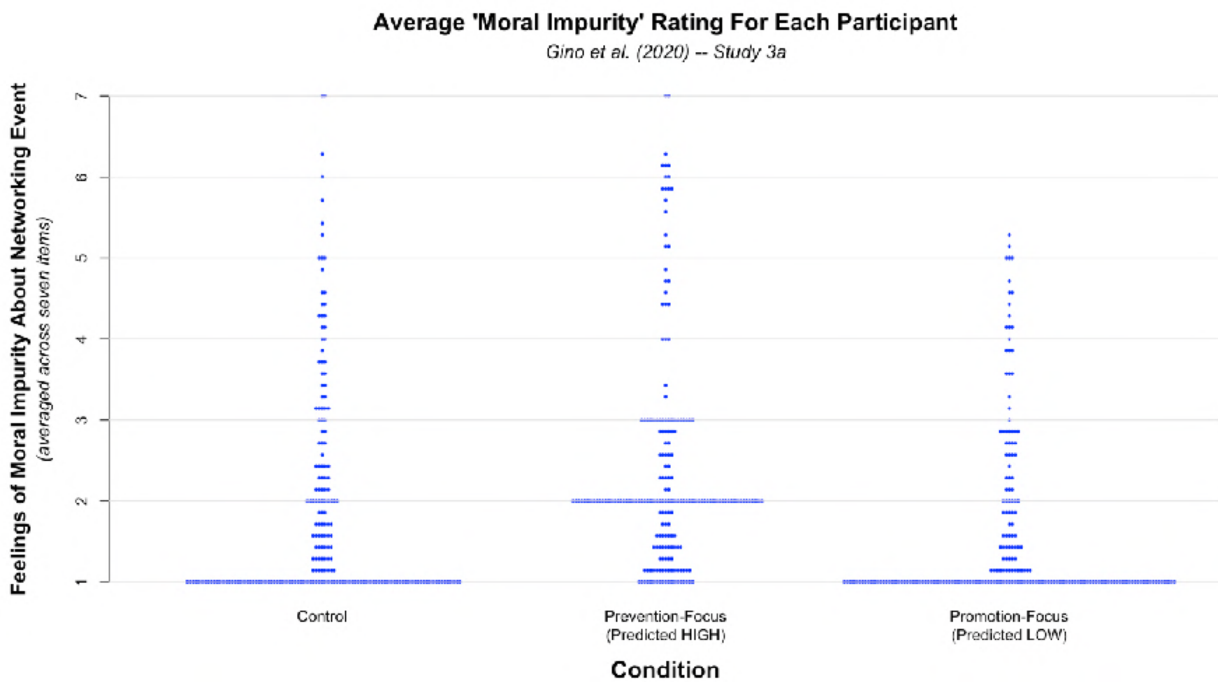
Screenshot of few rows of actual dataset for Study 3a

Let's walk through a few of the observations. The first row of data in the screenshot, corresponding to row 531 in the dataset, shows a participant who provided a '1' to all seven of the moral impurity items. This participant didn't feel *at all* dirty, inauthentic, impure, ashamed, wrong, unnatural, or tainted by imagining herself at the networking event. And indeed, if you look at the "words2_cond" column on the far right, you can see that what this participant wrote about the networking event - "socializing, party, impression, connections, work" – is perfectly <u>consistent</u> with those ratings. Her *ratings* were positive, and her *words* were positive. This makes sense.

The anomalies we discuss below pertain to rows in which participants' ratings and words are *inconsistent*, when either the ratings are negative and the words are positive, or the ratings are positive and the words are negative.

5.3.1 Many 2s and 3s

Keeping that in mind, let's look at all of the raw data from Study 3a, using the same kind of plot presented in the previous two sections. Each dot in the figure below represents the average moral impurity rating for a single participant.



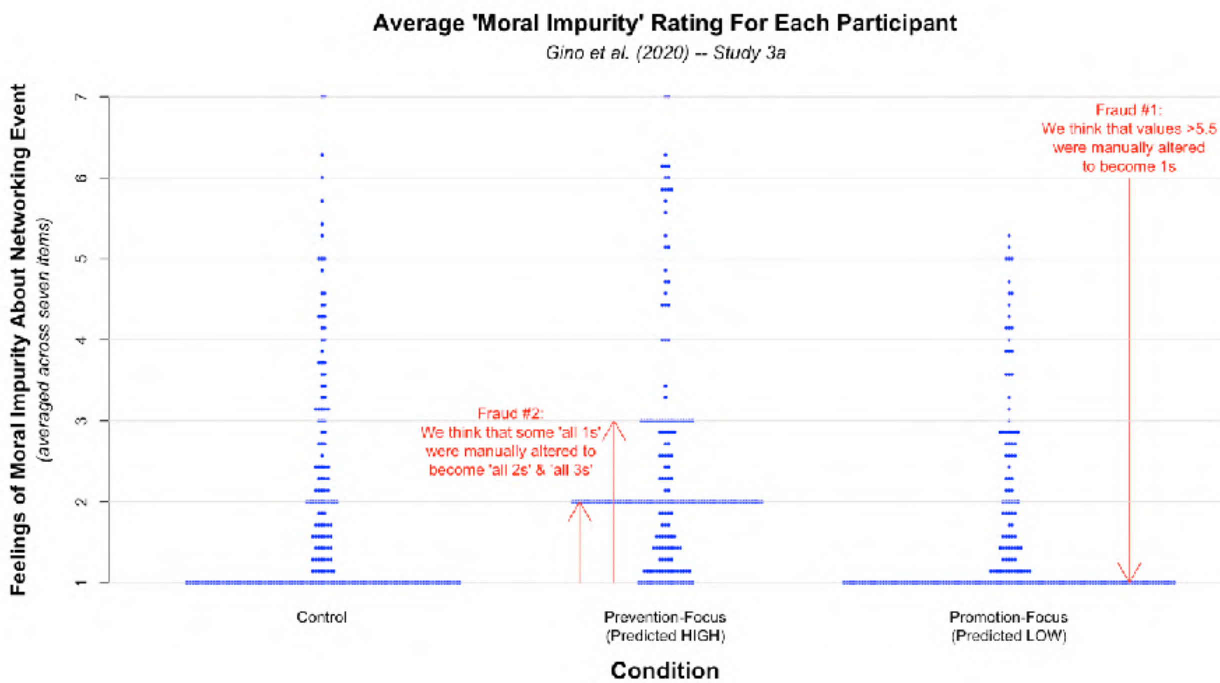
To start, consider the control condition, on the left. You can see that there are many participants with scores of 1.0, indicating that they did not feel *at all* dirty, inauthentic, impure, ashamed, wrong, unnatural, or tainted by imagining themselves at the networking event. We don't know how many 1.0s to expect, but it seems reasonable that many participants would wind up with this score. There is nothing intrinsically dirty about networking.

Now let's take a look at the dots in the middle, the prevention-focus condition. The authors hypothesized that writing a prevention-focused essay would *increase* participants' feelings of moral impurity when imagining the networking event. There is indeed a startling difference between the control condition and the prevention-focused condition: instead of '1.0' being the most common score on this dependent variable, now '2.0' is the most common score on this dependent variable. There is also a noticeable increase in the number of '3.0s.'

This is much more peculiar than it may seem at first. Remember that this dependent variable is an *average* of 7 items. There are obviously multiple ways for seven ratings to yield an average of 2.0 or 3.0, but the simplest and most common is for participants to give all '2s' or all '3s'. It is unusual for so many people to decide that they are across-the-board exactly a '2' on dirty, inauthentic, ashamed, etc. Indeed, ratings of 'all 2s' and 'all 3s' are quite rare in the other two conditions. In combination, the absence of '1.0s' and the presence of '2.0s' and '3.0s' led us to suspect that the researcher simply replaced many prevention-focused observations that were 'all 1s' with 'all 2s' or 'all 3s'. It is an easy way to tamper with the data. And it would of course yield the desired effect: higher moral impurity ratings among prevention-focused participants.

Keeping that in mind, let us turn to the promotion-focused condition on the right side of the figure. The authors hypothesized that writing a promotion-focused essay would *decrease* participants' feelings of moral impurity. And so here what we see is that there are *lots* of '1.0s', even more than in the control condition, accompanied by a complete absence of values greater than 5.5. That led us to suspect that the researchers replaced those high values with all 1s. Again, this would make the data tamperer' s job easy, and it would yield the desired effect, low moral impurity ratings among promotion-focused participants.

This annotated figure summarizes these two forms of hypothesized fraud:[8]



5.3.2 Participants with positive ratings and negative words (N=9)

Of critical importance here is the fact that participants both rated how morally impure they felt *and* wrote text describing how they felt, whereas the researchers cared *only* about the ratings (which they analyzed) and not about the text (which, therefore, they did not need to analyze). This means that a researcher who tampered with this data might have manually altered some participants' ratings *without also* feeling compelled to manually alter the text that accompanied those ratings. This would leave a trace. For those tampered observations, the valence implied by the ratings and the valence implied by the text would be *inconsistent.*

Let's walk through these two hypotheses. First, let's focus on the promotion-focus condition, for which we hypothesize that a researcher manually changed some very high values – values associated with extreme levels of moral impurity – into maximally low values – values associated with no moral impurity at all. If that is true, then we should see some participants in the dataset who (1) provided an average rating of 1.0 on the moral impurity scale *and* (2) wrote text suggesting that they felt extremely morally *im*pure. Moreover, those participants should be over-represented in the promotion-focus condition.

And, indeed, in this dataset we found nine participants who both averaged a 1.0 on the moral impurity scale *and* wrote text implying that they felt high levels of moral impurity. Of the nine, seven of them were in the promotion-focus condition:

[8] As emphasized in the previous section, we are not purporting to explain *entirely* what happened here, as it is possible that data tampering also took other forms in this study. We are merely suggesting that at least some of the data tampering was carried out in the way hypothesized here.

CumID_all	MI1	MI2	MI3	MI4	MI5	MI6	MI7	words2_cond	conditions
207	1	1	1	1	1	1	1	aggressive, pushy, calculating, egotistic, pushy	control
535	1	1	1	1	1	1	1	Wow, liar, false, delusional, braggart	control
118	1	1	1	1	1	1	1	I felt uncomfortable and inauthentic. The last thing I want to talk abo	promotion
248	1	1	1	1	1	1	1	Gross, phony, supercilious, unpleasant, disingenuous	promotion
335	1	1	1	1	1	1	1	Scummy; dishonest; disgusting; disingenuous; weak; unoriginal	promotion
359	1	1	1	1	1	1	1	All that corporate stuff is awful.	promotion
498	1	1	1	1	1	1	1	schmoozing, suck-up, ambition, networking, career, connections	promotion
538	1	1	1	1	1	1	1	dirty,fake,cheap,butt kisser,not a good person	promotion
589	1	1	1	1	1	1	1	gross slimy player suck up wrong	promotion

This is consistent with the notion that all or some of these apparent '1.0s' were not actually '1.0s'. The words they wrote suggest that they may have instead provided very high ratings on the moral impurity scale, ratings that were altered by the researcher performing the analysis.[9]

Though we find this evidence to be fairly convincing, it is not conclusive, as it suffers from the limitations of being somewhat subjective and also reliant on a small number of observations. The next analysis – which focuses on the hypothesis that some prevention-focused '1.0s' were manually altered to become '2.0s' and '3.0s' – does not suffer from either limitation.

5.3.2 Participants with negative ratings and positive words (N=79)

To perform *this* analysis, we relied on a technique known as "sentiment analysis," which uses an algorithm to score a passage of text on the dimension of valence. Using the VADER package in R, we used an algorithm that took in participants' textual description of the networking event, and gave it a score from 1 (maximum positivity) to -1 (maximum negativity). Essentially, the score reflects the net percentage of positive minus negative words in a text sample. If a string of text contains only unambiguously positive words, it will have a score of 100%, or 1.000; if it contains only unambiguously negative words, it will have a score of -100%, or -1.000. The screenshot below shows some participants whose VADER score was maximally positive (i.e., 1.000):

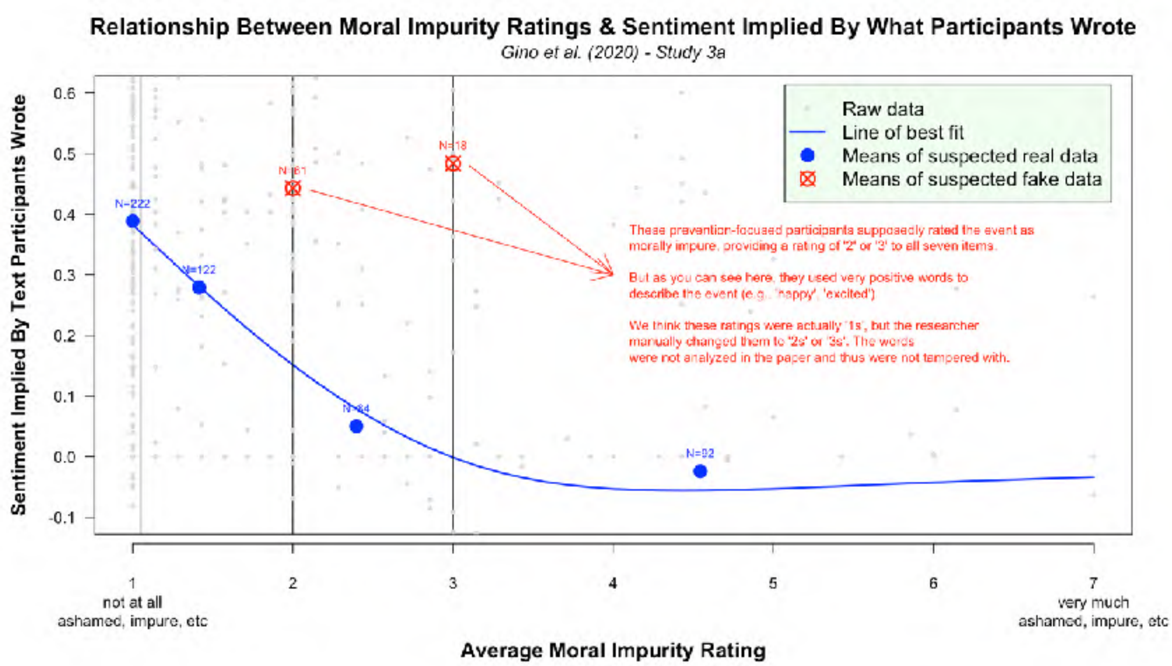
	s3a.net	s3a.words2_cond
18	1.000	Fun, confidence, honor, luck, privilege
30	1.000	excited, fun, hopeful, inspirational, strong, motivated
36	1.000	entertaining, exciting, fun, privileged, encouraging
44	1.000	excited, pleased, interested, smart, excited
115	1.000	Active, novel, proactive, ambitious, satisfied
160	1.000	Excited, focused, accomplished
192	1.000	Fun, excited, important
218	1.000	Happy; Smart; Euphoric; Intelligent; Joy; Celebrate
292	1.000	optimistic, happy
296	1.000	ambitious, determined, engaged, sociable, kind, smart
317	1.000	proud, worth, motivating, lucky, powerful
328	1.000	anticipation, excitement, happy, joy, pleasure
349	1.000	confident thrilled accomplished proud smart

And here are the participants with the most negative VADER scores:

[9] These are not the only '1.0s' who wrote somewhat negative things, but they were the only ones who wrote things implying moral impurity. For example, a few other '1.0s' mentioned feelings of anxiety or boredom.

	s3a.net	s3a.words2_cond
389	-0.810	gross, exhausting, tired, networking, yucj
371	-0.815	Sleazy, fake, disgusting, boring, pointless
284	-0.846	concerned, worried, angered
207	-0.891	aggressive, pushy, calculating, egotistic, pushy
351	-0.894	worried, stressed, trying, tough, confused
90	-0.903	fake, schmoozing, painful, awkward, weird
543	-0.915	fake, boring, exhausting, tiring, dreadful
399	-0.917	stressful, embarrassing, anxious, talking, fake
576	-0.928	cheating, disgusting, wrong, annoying, slimy
22	-1.000	Stressful, bad, anxiety, dislike, avoidance
68	-1.000	Repulsed, disgusted, tired, annoyed, irritated
305	-1.000	Uncomfortable
392	-1.000	bored, confused, unsure, uncertain, wtf

As indicated above, we believe that many of the ‘all 2s’ and ‘all 3s’ in the prevention-focus condition may actually have entered ‘all 1s’, and thus may have felt very positively toward the networking event. If this is true, and if, as we suspect, the researcher altered the moral impurity ratings without also altering the words those participants wrote about the networking event, then the words written by those ‘all 2s’ and ‘all 3s’ should look a lot like the words written by ‘all 1s’. They should be much too positive. The figure below is consistent with this prediction.



The blue line in this chart represents the observed relationship between the moral impurity ratings and the sentiment scores across all conditions, excluding the prevention-focused observations that we hypothesized to have been tampered with. The relationship is sensibly negative: More morally impure ratings are associated with lower sentiment scores and thus more negative text descriptions.

The two red dots with X’s depict the average sentiment scores of those in the prevention-focus condition who gave ratings of ‘all 2s’ and ‘all 3s’. If they were *really* ‘all 1s’ to begin with, then the text they wrote should be very positive, and thus their sentiment scores should be high. And that is exactly what we see

here. The 'all 2s' and 'all 3s' in the prevention-focused condition wrote text that was just as positive as what the 'all 1s' wrote across the entire sample. This very strongly suggests that a great many of these 'all 2s' and 'all 3s' were really 'all 1s' that had been altered.

6. Case #4: Study 4 of Gino & Wiltermuth (2014)

In this paper, the authors present five studies demonstrating that "dishonesty may lead to creativity".

Here we focus on Experiment 4, which was run online (using mTurk participants). We received this dataset from a researcher who had years ago obtained it from Professor Gino.

6.1 Procedure

In Experiment 4, 178 mTurk participants were first asked to guess whether the outcome of a virtual coin toss would be heads or tails. After indicating their prediction, participants had to press a button to toss the coin virtually. They were asked to press the button only once, but after that they were invited to press the button many times to make sure the coin was legitimate. This was designed to give participants room for justifying their own cheating. Participants reported whether they had guessed the coin toss outcome correctly, and they received a $1 bonus if they had. Because the computer recorded their predictions as well as the outcome of the coin toss, the experimenters could tell whether participants had cheated.

After completing a scale measuring rule-following (not discussed further in this report), participants completed two creativity tasks, a "uses" task and the Remote Associates Task.

Our analysis will focus exclusively on the results of the "uses" task, which involved asking participants "to generate as many creative uses for a newspaper as possible within 1 min" (p. 976).

6.2 Results

In line with the authors' hypothesis, participants who cheated on the coin toss task came up with more uses for a newspaper (M = 8.3) than did participants who did not cheat (M = 6.5), $p < .0001$.

6.3 Direct Evidence of Tampering

The dataset seems to be sorted by two columns, first by a column called "cheated", indicating whether participants cheated on the coin toss task (0 = did not cheat; 1 = cheated), and then by a column called "Numberofresponses", indicating how many uses for a newspaper the participant generated.

For example, the screenshot below depicts the first 40 observations in the dataset.[10] Because the data are sorted first by the "cheated" column, all of these observations represent non-cheaters (i.e., scores of 0 in that "cheated" column). The shown rows are perfectly sorted by the "Numberofresponses" column. Indeed, the 135 non-cheaters in the dataset are all sorted by the "Numberofresponses" column.

[10] To create this screenshot, we had to move the "cheated" and "Numberofresponses" columns. In the dataset that Gino shared, those variables were in the 78th and 14th columns, respectively.

1	StartDate	EndDate	MTurkID	Cum_ID	cheated	Numberofresponses
2	11/17/12 23:54	11/18/12 0:07	AD8VVYGP4LRKG	144	0	2
3	11/17/12 23:17	11/17/12 23:41	A2KJZAMH6G8LWC	91	0	2
4	11/17/12 23:44	11/17/12 23:57	A21TECY6SM7BNV	127	0	3
5	11/17/12 22:57	11/17/12 23:11	A2GRSJHXTR7JQR	24	0	3
6	11/18/12 0:00	11/18/12 0:20	A1FAQI6Q4WCS	168	0	3
7	11/17/12 23:41	11/17/12 23:52	A1YZJ7OO7Q2D89	113	0	3
8	11/17/12 23:37	11/17/12 23:47	AVA93G56VQLZA	101	0	3
9	11/17/12 23:20	11/17/12 23:32	A20863XUQJT5T1	76	0	3
10	11/18/12 0:11	11/18/12 0:24	A27I79P03I0ZPO	173	0	3
11	11/17/12 23:11	11/17/12 23:28	A12WY0ZDGVOZQS	69	0	3
12	11/17/12 23:41	11/17/12 23:56	A20552JTR91G67	124	0	3
13	11/17/12 23:17	11/17/12 23:33	A3Q9UUF8RPV4LQ	79	0	3
14	11/17/12 22:49	11/17/12 22:58	A2BH9W7Y1TL3X8	1	0	3
15	11/17/12 23:59	11/18/12 0:10	A034420738QHAX9TNO9BA	152	0	4
16	11/17/12 23:38	11/17/12 23:51	a32k7qy8nwzx43	110	0	4
17	11/17/12 23:05	11/17/12 23:23	A2DAT0DBUXU8FF	55	0	4
18	11/17/12 23:39	11/17/12 23:49	A20A0EM29IJLSK	103	0	4
19	11/17/12 23:31	11/17/12 23:51	APJEYYRENCAC6	109	0	4
20	11/17/12 23:02	11/17/12 23:27	A1L6EDKEUG69XB	66	0	4
21	11/18/12 0:00	11/18/12 0:10	AYZ00GXI5D15Y	150	0	4
22	11/17/12 23:22	11/17/12 23:35	APHNYDGTCRN3O	82	0	4
23	11/17/12 23:19	11/17/12 23:32	A1MM8TSLCHVMNK	75	0	4
24	11/17/12 23:12	11/17/12 23:24	A3AZJGI9D7C0PD	57	0	4
25	11/17/12 22:52	11/17/12 23:17	A3DQUF5TM9VTS7	37	0	4
26	11/17/12 23:50	11/18/12 0:03	A77M840AXJ16B	137	0	4
27	11/18/12 0:02	11/18/12 0:10	A3GSCVUHX7DM8T	151	0	4
28	11/17/12 23:05	11/17/12 23:24	A26L91YL0GDGD8	58	0	4
29	11/17/12 23:27	11/17/12 23:53	AJY9CIX7FW9W1	115	0	4
30	11/17/12 23:48	11/18/12 0:02	ALSE4C4Q3R6G	133	0	5
31	11/17/12 22:54	11/17/12 23:08	A2R8SVW42IYFYX	17	0	5
32	11/17/12 22:59	11/17/12 23:17	A07109741WN0LPDUN9GL9	34	0	5
33	11/17/12 23:25	11/17/12 23:37	A1GFD4B3NOMWIY	86	0	5
34	11/17/12 23:37	11/17/12 23:54	ADQML8ECWYME5	119	0	5
35	11/17/12 23:04	11/17/12 23:32	A3FAAKASDY5HE6	183	0	5
36	11/17/12 22:55	11/17/12 23:14	A5SUR5C68YYN8	30	0	5
37	11/17/12 22:56	11/17/12 23:08	A2MBAN2GDK1P1J	16	0	5
38	11/17/12 23:48	11/18/12 0:00	A34N9G0IEI28IG	131	0	5
39	11/17/12 23:46	11/18/12 0:06	A3AHNUDEOZ33JE	143	0	5
40	11/17/12 23:25	11/17/12 23:38	A1QK6O24KDVLJ1	88	0	5
41	11/17/12 22:58	11/17/12 23:19	A7NLUN5YH4S9L	43	0	5

The next screenshot, in contrast, shows that while 43 cheaters are also sorted by this variable, there are 13 observations that are not in the order they should be.

1	StartDate	EndDate	MTurkID	Cum_ID	cheated	Numberofresponses
132	11/18/12 0:04	11/18/12 0:13	A1X82CGYFM586F	155	0	11
133	11/17/12 23:08	11/17/12 23:22	A1F14BB4PV053A	53	0	11
134	11/17/12 23:22	11/17/12 23:37	A356ZZWYC8GRVY	85	0	11
135	11/17/12 23:44	11/18/12 1:05	A34DG3IZ88WWBT	192	0	12
136	11/17/12 22:58	11/17/12 23:14	A3P7XKTEBOKN5R	29	0	13
137	11/18/12 0:01	11/18/12 0:20	ADTNOFJHTTB1L	167	1	3
138	11/17/12 23:34	11/17/12 23:53	A1UNAJF3E5HH17	114	1	3
139	11/17/12 23:44	11/17/12 23:57	A0377367199XXE56OT9GZ	126	1	4
140	11/17/12 23:36	11/17/12 23:46	A2DUKWR9I6FFZV	99	1	4
141	11/17/12 23:02	11/17/12 23:17	AE3D6SE2D8UPQ	36	1	13
142	11/17/12 23:32	11/17/12 23:43	A21MCWTDIKATV5	97	1	9
143	11/17/12 23:59	11/18/12 0:10	A28XLOE0FMG1ZX	153	1	5
144	11/17/12 22:55	11/17/12 23:04	A126XP3VIWJKD6	8	1	5
145	11/18/12 0:07	11/18/12 0:21	A3E1EPRY1OYE34	171	1	9
146	11/17/12 23:30	11/18/12 0:03	A27AEIRFEFR4US	136	1	5
147	11/17/12 23:30	11/18/12 0:44	A07854333QXCSICF01THG	191	1	9
148	11/17/12 23:38	11/17/12 23:50	A311BZDLCK6HQQ	105	1	8
149	11/17/12 22:59	11/17/12 23:15	A1ILA0RGDB9JJ6	32	1	9
150	11/17/12 23:11	11/17/12 23:22	A22LZ62E0UC4VL	51	1	5
151	11/17/12 23:49	11/18/12 0:03	A1SHH0U3JHSCSV	187	1	6
152	11/18/12 0:03	11/18/12 0:22	A37JD0XUZHQYRC	172	1	6
153	11/17/12 22:52	11/17/12 23:04	ALML8V38FDV0	9	1	9
154	11/17/12 23:58	11/18/12 0:14	A3W4736CCV8TT4	157	1	11
155	11/18/12 0:07	11/18/12 0:15	AUN8AE8UCO3MD	159	1	14
156	11/17/12 23:13	11/17/12 23:29	Jazzy67033	180	1	6
157	11/17/12 22:58	11/17/12 23:08	A208MTGA7V29TP	14	1	8
158	11/17/12 23:51	11/18/12 0:08	A2UL07RCD2RO8R	146	1	10
159	11/17/12 22:51	11/17/12 23:10	AP37A6DG5TTEM	20	1	7
160	11/18/12 0:03	11/18/12 0:14	A2H18EYM79ZRCW	156	1	7
161	11/17/12 23:59	11/18/12 0:09	A1BCCFEEN32OWP	149	1	8
162	11/17/12 23:03	11/17/12 23:15	A3TN3GQAO618VB	31	1	7
163	11/18/12 0:03	11/18/12 0:21	hhendric@hotmail.com	169	1	7
164	11/17/12 23:13	11/17/12 23:26	A62RZY5BWOZZM	63	1	14
165	11/17/12 23:25	11/17/12 23:47	AVUAN8WKJ443M	102	1	8
166	11/17/12 23:48	11/17/12 23:59	A25KU26Y8FTJPV	129	1	8
167	11/17/12 22:55	11/17/12 23:06	A3OF0DCN3KU8HT	11	1	8
168	11/17/12 23:11	11/17/12 23:18	A47QHTQNUTOVL	42	1	8
169	11/17/12 23:52	11/18/12 0:03	A1ASPIEIOZXL3U	138	1	8
170	11/17/12 23:27	11/17/12 23:32	A3E0AY1XXP8IBQ	77	1	9
171	11/17/12 23:57	11/18/12 0:21	A1R7CJMWXC79UO	170	1	10
172	11/17/12 23:03	11/17/12 23:10	A5VWAZZ49D5WU	22	1	10
173	11/17/12 23:21	11/17/12 23:31	AGX6FRHVVU2WS	74	1	10
174	11/17/12 23:25	11/17/12 23:37	A24JC2CF7MMG41	84	1	10
175	11/17/12 23:46	11/17/12 23:58	A1REWUVT3N8SN7	128	1	11
176	11/17/12 23:37	11/17/12 23:50	A27MJOV91GA8R3	106	1	11
177	11/17/12 23:06	11/17/12 23:17	A17M7G85OEIB3U	35	1	11
178	11/17/12 23:06	11/17/12 23:21	A2IF1VIC7GZUN	50	1	12
179	11/17/12 23:07	11/17/12 23:17	A2GPIQQ2PJ87Q0	38	1	13

As was the case with previous datasets, we believe that these observations were manually altered to produce the desired effect.

There are three things worthy of note here.

First, as before, it is not possible to sort the dataset to generate the order in which the data were saved. They were either originally entered this way (which is implausible, since the data originate in a Qualtrics file, which by default sorts by time), or they were manually altered.

Second, because rows are sorted by the variable of interest, "numberOfUses", if the values that are out of order were changed, it is straightforward to impute what they were changed from. For example, row #141 is "13", the number right before it is "4", and the first non-suspicious value after it is "5". Therefore, if the data were changed, then we can assume that that "13" used to be either a "4" or a "5".

One can do this for each of the 13 highlighted values in the dataset. We can thus reconstruct what the data looked like before they were tampered with. The screenshot below shows the imputed values for all relevant cells. The first new column ("Imputed1") imputes the lowest value that is consistent with the neighboring observations, and the second new column ("Imputed2") shows the highest value. So we see, for example, that that first "13" could have been either a "4" or a "5".

1	StartDate	EndDate	MTurkID	Cum_ID	cheated	Numberofresponses	Imputed1	Imputed2
137	11/18/12 0:01	11/18/12 0:20	ADTNOFJHTTB1L	167	1	3	3	3
138	11/17/12 23:34	11/17/12 23:53	A1UNAJF3E5HH17	114	1	3	3	3
139	11/17/12 23:44	11/17/12 23:57	A0377367199XXE56OT9GZ	126	1	4	4	4
140	11/17/12 23:36	11/17/12 23:46	A2DUKWR9I6FFZV	99	1	4	4	4
141	11/17/12 23:02	11/17/12 23:17	AE3D6SE2D8UPQ	36	1	13	4	5
142	11/17/12 23:32	11/17/12 23:43	A21MCWTDIKATV5	97	1	9	4	5
143	11/17/12 23:59	11/18/12 0:10	A28XLOE0FMG1ZX	153	1	5	5	5
144	11/17/12 22:55	11/17/12 23:04	A126XP3VIWJKD6	8	1	5	5	5
145	11/18/12 0:07	11/18/12 0:21	A3E1EPRY1OYE34	171	1	9	5	5
146	11/17/12 23:30	11/18/12 0:03	A27AEIRFEFR4US	136	1	5	5	5
147	11/17/12 23:30	11/18/12 0:44	A07854333QXCSICF01THG	191	1	9	5	5
148	11/17/12 23:38	11/17/12 23:50	A311BZDLCK6HQQ	105	1	8	5	5
149	11/17/12 22:59	11/17/12 23:15	A1ILA0RGDB9JJ6	32	1	9	5	5
150	11/17/12 23:11	11/17/12 23:22	A22LZ62E0UC4VL	51	1	5	5	5
151	11/17/12 23:49	11/18/12 0:03	A1SHH0U3JH5CSV	187	1	6	6	6
152	11/18/12 0:03	11/18/12 0:22	A37JD0XUZHQYRC	172	1	6	6	6
153	11/17/12 22:52	11/17/12 23:04	ALML8V38FDV0	9	1	9	6	6
154	11/17/12 23:58	11/18/12 0:14	A3W4736CCV8TT4	157	1	11	6	6
155	11/18/12 0:07	11/18/12 0:15	AUN8AE8UCO3MD	159	1	14	6	6
156	11/17/12 23:13	11/17/12 23:29	Jazzy67033	180	1	6	6	6
157	11/17/12 22:58	11/17/12 23:08	A208MTGA7V29TP	14	1	8	6	7
158	11/17/12 23:51	11/18/12 0:08	A2UL07RCD2RO8R	146	1	10	6	7
159	11/17/12 22:51	11/17/12 23:10	AP37A6DG5TTEM	20	1	7	7	7
160	11/18/12 0:03	11/18/12 0:14	A2H18EYM79ZRCW	156	1	7	7	7
161	11/17/12 23:59	11/18/12 0:09	A1BCCFEEN32OWP	149	1	8	7	7
162	11/17/12 23:03	11/17/12 23:15	A3TN3GQAO61BVB	31	1	7	7	7
163	11/18/12 0:03	11/18/12 0:21	hhendric@hotmail.com	169	1	7	7	7
164	11/17/12 23:13	11/17/12 23:26	A62RZY5BWOZZM	63	1	14	7	8
165	11/17/12 23:25	11/17/12 23:47	AVUAN8WKJ443M	102	1	8	8	8
166	11/17/12 23:48	11/17/12 23:59	A25KU26Y8FTJPV	129	1	8	8	8
167	11/17/12 22:55	11/17/12 23:06	A3OF0DCN3KU8HT	11	1	8	8	8

Third, when one reconstructs the data in this way, by replacing the highlighted values with the values one would impute based on the order in which data are sorted, the significant relationship between cheating and creativity on the uses task entirely disappears. It's p-value goes from <.0001 to .292 ("Imputed1") or .180 ("Imputed2").

7. Reminder

This report includes a subset of the evidence of tampering we have collected, which was obtained by analyzing a small subset of the data that Gino has published.